WHO consultation on Data and Results Sharing During Public Health Emergencies.

Background Briefing.

Sep 2015

Background Briefing for WHO consultation on Data and Results Sharing During Public Health Emergencies.

Authors: Ben Goldacre, Sian Harrison, Kamal R. Mahtani and Carl Heneghan

Corresponding author
ben.goldacre@phc.ox.ac.uk

Centre for Evidence-Based Medicine,
Nuffield Department of Primary Care Health Sciences,
University of Oxford

ACKNOWLEDGEMENTS

SUMMARY

Background
This report was commissioned as a short background briefing for a WHO consultation meeting. CEBM was asked to: interview a range of stakeholders provided by WHO; develop a set of themes from these interviews; and supplement these themes where possible with a rapid literature review on the issues raised.

The recent Ebola outbreak has raised important issues around rapidly sharing data and results in public health emergencies. Rapid data sharing is important to help identify the causative agent; investigate and predict disease spread; define diagnostic criteria; and evaluate treatments and methods to contain further spread.[1] The types of information being shared include surveillance data, trial data, pathogen genome data, case reports, and then summary results derived from these data sources. The consumers of this data and research include public health officials, clinicians, aid workers, researchers, governments and NGOs, who use this information to develop and implement treatments for individuals, and to bring epidemics under control.

Multiple barriers to rapid information sharing were identified.

The most commonly raised concern was around *data protection* and the confidentiality of potentially re-identifiable data on individuals. This includes ethical concerns about whether *consent* has been given and whether it covers secondary uses; and barriers from individual countries' data protection *legislation*. There is a long history of such data being shared in epidemiological research, and an extensive knowledge base of strategies to mitigate and manage risk.

The tension between rapid sharing and accuracy was often cited, with examples of harm from early release of incorrect information in previous emergencies. Mitigation strategies here include "buyer beware" caution on early release, and extensive metadata on provenance of early data for subsequent checking.

Political and cultural issues raised included countries witholding information through a well-founded concern about negative practical consequences of wider awareness of outbreaks, such as loss of trade.

Reciprocity was raised in two contexts. *Nations* may be concerned that data they have collected and shared will be used to produce treatments they cannot afford. *Individuals* may be concerned that data they have collected and shared will be used to produce academic papers without recognising their input. This in turn can have a negative impact on sustainability for research capacity and data collection within countries.

Poor systems for knowledge curation were raised, with a concern that these issues had often been discussed in the abstract without adequate concrete progress.

Issues around academic journal publication were raised, specifically the delays and extra workload imposed through this means of knowledge dissemination. This was borne out by qualitative and quantitative research on the same topic demonstrating extensive non-publication and publication delay on research conducted during public health emergencies. Related challenges were identified for *genome data sharing*, although this poses fewer challenges and is often done well.

Lastly there were concerns around *lack of time and resources* for adequate dissemination and communication.

Multiple solutions and opportunities were identified

Interviewees raised general issues around the process of developing solutions. This included the need to *learn from other fields, and previous experiences*; the need to *create enduring systems and cultures* for data sharing, since the cast of players can change with each new emergency; the need to *engage and listen to the right people*, whether organisations or individuals, to ensure that needs and misgivings are identified and addressed; the need for *clear positive commitments* from individuals and organisations in influential positions; and the need for *capacity building* around sharing data and knowledge so that this can be addressed as a legitimate foreground topic for working academics and others, rather than an afterthought.

There was no single clear, simple, immediate solution identified.

Regulatory frameworks that are consistent, clear, and versatile were repeatedly recognised as important.

Various opportunities around *knowledge curation* were discussed, including *formal data sharing platforms* and the value of other forms of *curation, synthesis and dissemination*. A range of existing platforms are briefly reviewed, including WWARN, GISAID, EbolaClinicalTrials, ISARIC, and specific genome data projects.

Multiple opportunities for *improving journal publication* were identified, including re-asserting existing commitments on rapid review and protocol acceptance; and new publishing models such as pre-print servers and post-publication peer review.

There was a view that *academic reward structures* could be modified to reflect the value of data sharing.

In addition, this report's authors have set out some brief additional suggestions including pilots of "ideal" regulatory frameworks and knowledge curation models; ongoing live audit of results dissemination from completed research; sharing peer reviews when a journal article is rejected, to expedite review in a subsequent journal; and legally robust but readily understandable off-the-shelf legal agreements for data sharing, to address concerns around reciprocity, modelled on similar projects from other fields.

KEY MESSAGES

Problems:

1. Sharing data and knowledge is extremely important to save lives.

2. Current initiatives have not been sufficient to deliver rapid sharing.

3. Some reluctance to share data is well founded, and reflects a desire for reciprocity which may carry a financial or time burden for those invited to reciprocate.

4. The rate of publication of completed studies from emergency settings is extremely poor.

Solutions

1. Improving data sharing will require concrete initiatives and resources, alongside culture change.

2. These will need to be developed now, and before - rather than during - the next emergency.

3. There is no single magic bullet.

4. There are clear opportunities with journals, better curation, rewards, and standardising agreements

5. Challenges around sharing re-identifiable data on individuals have been overcome in other settings.

Background of the report

This brief report was commissioned by WHO from CEBM at short notice as a rapid background briefing to a WHO consultation meeting. CEBM were asked to: interview a range of stakeholders provided by WHO; develop a set of themes from these interviews; and supplement these themes where possible with a rapid literature review on the issues raised.

Methods:
A series of semi-structured interviews were conducted with key individuals selected and identified by the WHO as capable of informing this discussion.

Three authors (SH, BG, CH) agreed a structure for the interviews which included two main lead-in themes for the main interviewer (SH) to follow: problems with rapid data and results sharing during public health emergencies; and potential solutions to these problems whether past, present or future. The interview schedule is available on request and was emailed to most interviewees in advance. The telephone interviews were recorded and then transcribed. Results of the interviews were reviewed and grouped into themes through an iterative process until agreement among the authors was reached. The results of the interviews were supplemented with a rapid overview of relevant current literature.

Results:
A total of 13 interviews took place with individuals based in North America, Europe and Africa. One interview was conducted via email, the rest via telephone. Verbal consent was sought and given from all interviewees. The interviews lasted between 10 and 55 minutes. All quotes are currently anonymised.

Limitations:
This was a low resource review, interviewing a small number of researchers, and conducting a brief non-systematic review of the literature, on a tight timescale.

1. Problems with sharing data and results in public health emergencies:

The importance of competent and rapid data sharing, and the presence of ongoing challenges, was universally recognised by all participants. A range of themes emerged around barriers to better sharing including: data protection; accuracy of data; governmental and political challenges; personal motivations of researchers; poor systems for sharing; problems with academic journal publication; and lack of time and resource in emergency settings. These are covered in more detail below.

1.1 Data protection

Concern over the confidentiality of data about individuals was the single most consistently cited barrier to data sharing. Participants described the conflict between the need to protect the participants' privacy, and the need for expediency to benefit the population at large.

- *'Part of the challenge is balancing the need for keeping an individual patient's data private but at the same time maximizing the insights from it so that other clinicians that are trying to work in this area and other patients that are sick can benefit from any knowledge learned'*
- *'There is some risk of identification. On the other hand, the public need to know can sometimes outweigh that'.*

This is a well recognised issue in data sharing. Reidentification of individuals can be particularly problematic if it leads to stigmatisation or other harms, and such risks may vary with cultural contexts. The same issue also emerged as a recurring theme in a range of quantitative and qualitative research projects published this month (Aug 2015) in a special issue of the Journal of Empirical Research on Human Research Ethics.[2] It is also a well recognised problem in epidemiology more broadly, where there is longstanding experience of sharing identifiable and potentially re-identifiable datasets for observational research (e.g. paitents' electronic health records) and individual participant data from randomised trials for Individual Patient Data (IPD) metaanalysis (the first of which was published in 1970[3]).

The standard approaches to managing the information security risks of IPD sharing are set out in the IOM report[4] on clinical trial data sharing, and elsewhere. They include: sharing IPD only to trusted individuals; assessing and augmenting information security capability of the recipient; a secure legal framework and penalties for data security breaches; improving pseudonymisation, for example by sharing minimum necessary fields for the work, while recognising that data can often potentially be re-identified through linkage to other datasets, or through the rarity of the incidence; auditing data usage; IT solutions such as "sandboxing" to permit analysis for summary results without downloading a full dataset; and watermarking data releases so that leaks can be traced.

Consent and governance were also raised. There is considerable disagreement in previous commentary and qualitative research about whether specific consent is required for all secondary uses of individual patient data, beyond the use for which consent was initially sought, or where consent was never sought.[5] Seeking repeat consent for secondary use is not without risk: it can be burdensome, expensive, impossible, or expose an individual to stigmatisation by being identified as of medical interest. It is broadly but not universally agreed that for data collection from now on, initial consent should also ideally include consent for subsequent secondary re-uses of the data.

Legal frameworks, such as data protection laws, may act as a barrier to data sharing within or between countries. The legality of data sharing of public health data and access to health information without consent in a public health emergency setting has been tested in court, including the US Supreme Courts[6]. A full review of the legal issues is beyond the scope of this briefing, but legislation

often does not permit exceptions on public health grounds. There are some international legal frameworks and treaties that directly address certain specific issues, such as the Nagoya Protocol of the convention on biological diversity, and the biological weapons convention. A recent Chatham House publication on Overcoming Barriers to Data Sharing in Public Health, from a series of expert interviews, concluded that a global legal framework to address all issues around data sharing in public health emergencies is unlikely to be successful, and recommends a 'global data governance or ethical framework, supplemented by local memoranda of understanding that take into account the local context.' [7]

1.2 Accuracy of data

Interviewees consistently identified the risk of misinformation, and a tension between rapid sharing, and accurate information.

- *'You want things that are confirmed before you jump to sharing, you need reasonably corroborated confirmed information or it can lead to the wrong guidelines'*
- *'There's a real danger in releasing results you haven't got confidence in'*
- *'Despite the need for rapid sharing of data, caution must be deployed when presenting the data. If data is preliminary or based on underpowered analyses, this must be highlighted.'*

The tension between timelliness and accuracy is recognised in the literature and borne out by previous experience. Early on in the German 2011 E. coli O104 outbreak, a local health authority released preliminary data suggesting the source was cucumbers from Spain. This resulted in some countries suspending some Spanish food imports, although the outbreak was ultimately traced to sprouted fenugreek seeds imported from Egypt and sprouted in Germany. The economic costs were not trivial, and overall this outbreak resulted in losses to EU farmers of over €800 million. [8]

Various mitigation strategies have been proposed, beyond simply encouraging researchers, health services, governments, the public, and the media to be cautious about preliminary data. For example, MapAction is a humanitarian mapping charity providing geodata services in emergencies, and emphasised the importance of good metadata to track the provenance of data shared during the Ebola outbreak.[9] This can facilitate review and further checking against primary sources.

The IOM report on responsible data sharing also highlights the need to safeguard against invalid *interpretations* of data, for example with secondary analyses of trial data 'which could undermine trust in clinical trials or otherwise harm public health.' [4] This must be balanced against the need to subject primary analyses, which can themselves be flawed, to critical review.

1.3 Political and Cultural issues

Several interviewees raised the problem of countries not releasing data due to potential negative impacts on the source country and its citizens, industries, government, or reputation.

- *'I come from a country where lots of things were pushed under the carpet: don't tell about outbreaks, maybe they'll never know and, if they don't know, then it didn't happen'*
- *'There may be concerns that the data could be damaging to a country's reputation or to trade and tourism, or when talking about diseases that are highly [stigmatised].'*

Such issues can arise due to fears over restrictions on outward migration, or discouraging tourism and trade, [10] as well as worries about the economic or political implications of a public health

emergency. [11] It has been reported that obstacles to information sharing contributed to delays in responding to SARS CoV in 2003, and similarly to delays in responding to cholera outbreaks.[12,13]

Although these disadvantages must be balanced against wider health needs, concerns about practical consequences of data sharing are often well-founded. Travel restrictions are imposed during public health emergencies. Trade limitations imposed on the UK following cases of Bovine Spongiform Encephalopathy and their link to variant CJD resulted in the loss of billions of pounds in trade,[14] and the SARS outbreak is estimated to have cost $40-80bn, with costs especially borne by tourism industries[15]. The 2005 revised International Heath Regulations (IHR) [16] was intended, in part, to address this problem of countries not reporting the presence of infectious diseases through concern about economic consequences.

1.4 Intellectual Property and reciprocity to nations

Some interviewees raised the issue of fair recompense for countries sharing data, and whether individual countries' views and expectations are taken into account.

- *'Policies driven by the Western world don't take into account the perspectives of people from other countries, particularly from low and middle income countries.*
- *'Concerns range from use of the data against the government, to having the data not benefit the population, to asking for financial incentive to conduct the study or collect the data... This isn't universally the case, but in my experience is related to the fragility of the government and location that the crisis occurs.'*
- *'Reciprocity of data sharing relationships must be established.'*

Intellectual property concerns are also recognised in the literature as a potential barrier to sharing, or source of delay[17], alongside the notion that data sharing agreements can lead to exploitation or inequity. As an example: in 2006 Indonesia claimed 'viral sovereignty' over samples of H5N1 collected within its borders, citing concern that they would be used to produce expensive treatments or vaccines that the country itself could not afford. As a consequence these samples were not released to the WHO until an equitable means of sharing the benefits was negotiated. These negotiations, which led to the Pandemic Influenza Preparedness Framework (the PIP) took almost 5 years, a significant delay. Data sharing may therefore need to be linked to reciprocally beneficial intellectual property agreements that promote both sample and data sharing. However, the negotiation processes are not simple and take time.

1.5 Personal motivations and reciprocity to individuals

Several interviewees raised a related issue: of individual researchers, research groups, and countries being reluctant to share data, through a desire to exploit it for their own professional advancement. Several interviewees expressed the view that concerns around data ownership and privacy can sometimes be used as an "excuse" for not sharing results, where the underlying reason was actually personal gain through academic publication.

- *'I think people wanted to hold on to the data so that they could publish papers themselves about what happened and they didn't want to share it. It's often couched behind the issue of privacy'*
- *'They really haven't had a chance to dig deeply into the data and they feel they should have that opportunity'*
- *'There is a lack of reciprocity in data sharing agreements with certain members of the consortia feeling exploited and being used as sources of data but not recipients'*

The issue of individuals wishing to reap the benefit of datasets they have themselves created, through investing time and resource, is well recognised in the broader literature on academic data sharing outside of emergency settings. It is a well-founded concern, as analysis and academic publication are a source of recognition and promotion, which in turn are used to prioritise funding applications, and so recognition of effort and output by these primary researchers may be essential for data collection itself to be sustainable.

Interviewees raised this as an issue for researchers in high income countries, and low and middle income countries. In LMIC concern about reciprocity also raises the issue of capacity building[18]. Some LMIC researchers may lack the resources or training to conduct analyses of raw data rapidly, or to high standards, despite having invested significant time and resource in generating that data[19]. They have a legitimate interest in recognition through, for example, co-authorship, but also in capacity building, through more balanced and meaningful collaboration and training. This potentially offers global benefits, by creating independent academic researchers with a wider range of backgrounds and good knowledge of local datasets, culture, opportunities, politics, and health issues. There have consequently been calls for "fair trade, rather than free trade, in data".[20]

1.6 Poor systems for sharing and curation

The difficulty of poor communication between different research groups, organisations, regions and countries was repeatedly highlighted. Several interviewees referred to shortcomings of systems for knowledge curation and data sharing. Often circulation of information was left to informal social networks. A specific concern was also raised of permissions for sharing being set very strictly at the outset, resulting in unnecessarily restricted circulation for crucial information much later. The move from a default of not sharing data to a default of sharing data should be encouraged.

- *'A coordinated approach is lacking. Currently, attempts to share data in public health emergencies is ad hoc and haphazard, there is no structure set out in advance'*
- *'It's difficult if you are not in consortia to get information about what is going on currently'*
- *'In terms of getting information from outbreak sites, this is problematic and it's difficult to obtain this information, not because people are bad willing, but simply because it's difficult to get in touch'*
- *'There are challenges even within consortia who are working to answer the same question or to solve the same problems'*

A range of initiatives for data and results sharing were raised, but many interviewees felt that, while there had been extensive discussion of policy and ideas, and some platforms to address specific smaller and more tractable data sharing needs, the broader problem of rapid dissemination and knowledge curation had not been adequately addressed. Lack of standardization across national surveillance schemes often makes comparisons between countries impossible. For example, different criteria exist for simple diagnostic definitions such as influenza-like illness. Research collaboration hubs such as ISARIC (below) have therefore attempted to set precedents for standardizing data requirements and subsequent data sharing.
A brief overview of some of these specific initiatives is presented below.

Lastly, there was also concern that sometimes information was left unshared, for reasons that were never publicly elucidated.

- *'Trying to look at new therapeutics for treating Ebola, one of the challenges we ran into was that the patients, particularly in the Western world, that were being treated, the data from their treatment was not being released to anybody else to be able to learn whether the therapeutics were effective at all and so it meant that people had to keep trying things in the dark'*

1.7 Issues with Journal Publication

Participants identified a conflict between rigorous and lengthy peer review for publication in high impact journals, and the need for rapid dissemination of information. Some discussed fast tracking of publications during public health emergencies, others weren't aware of any such initiatives or improvement. Some interviewees also discussed the tension between making preliminary pragmatic reports available, and producing reports to the time-consuming high standards of an academic publication.

- *'We certainly haven't felt that the journals have particularly gone out of their way to fast track the process to making the results available'*
- *'To get those papers early on when their interest is the greatest we've had to ask the investigators to stop what they were doing - because quite often they were also caregivers for the patients suffering in the epidemic - and to write it up.'*

This issue is reflected in both quantitative and qualitative literature on results sharing in public health emergency settings. Data sharing requirements from mainstream medicine, even when adhered to, are likely to be inadequate in rapidly evolving public health emergencies. Quantitative research on timely and complete publication in public health emergencies shows that most research is published after the epidemic is over. Analysis of over 300 journal articles from the 2003 SARs outbreak found that only 22% of studies were submitted during the outbreak, with 8% accepted and 7% published during this time (figure below).[21] Even though submission-to-acceptance and acceptance-to-publication intervals were shorter than non-SARs articles, these intervals remained extremely long.
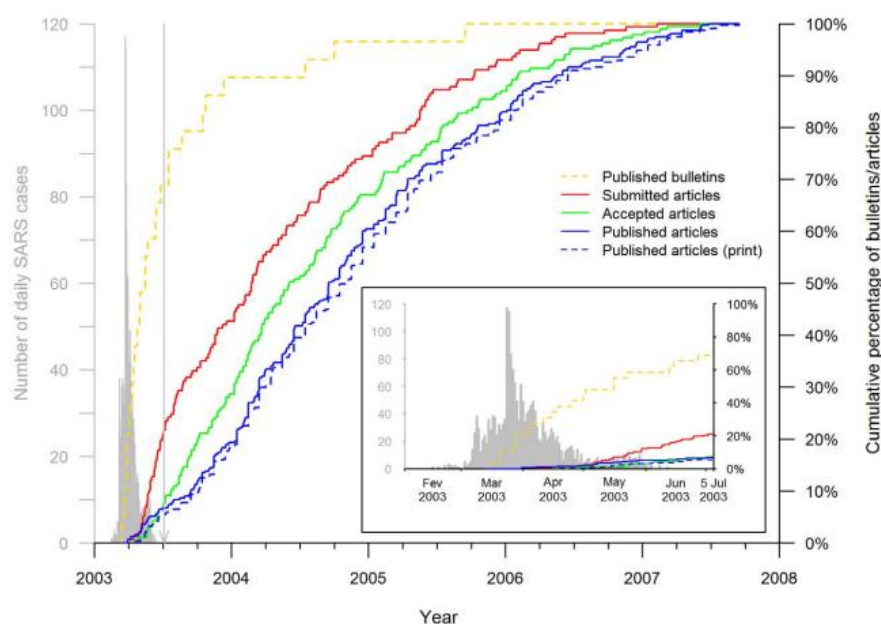

Figure: Submission/publication of SARS epidemiology papers, alongside number of SARS cases (grey), Xing et al, 2010.

Understanding of past pandemics may also be flawed, as information is likely to be incomplete. An observational survey of the publication delay of randomized trials on 2009 influenza A (H1N1) vaccination found 73 eligible trials registered in 2009-10, completed within a median of 5 months from their start date, but only 21 (29%) were published within one year of completion, and two years after the pandemic still only a minority had been published (figure below)[22].

This is concerning, as preventive strategies are largely if not wholly based on previous experience. Some of the components of current pandemic influenza plans were first developed and used during the Spanish Influenza pandemic of 1918–1919. [23]
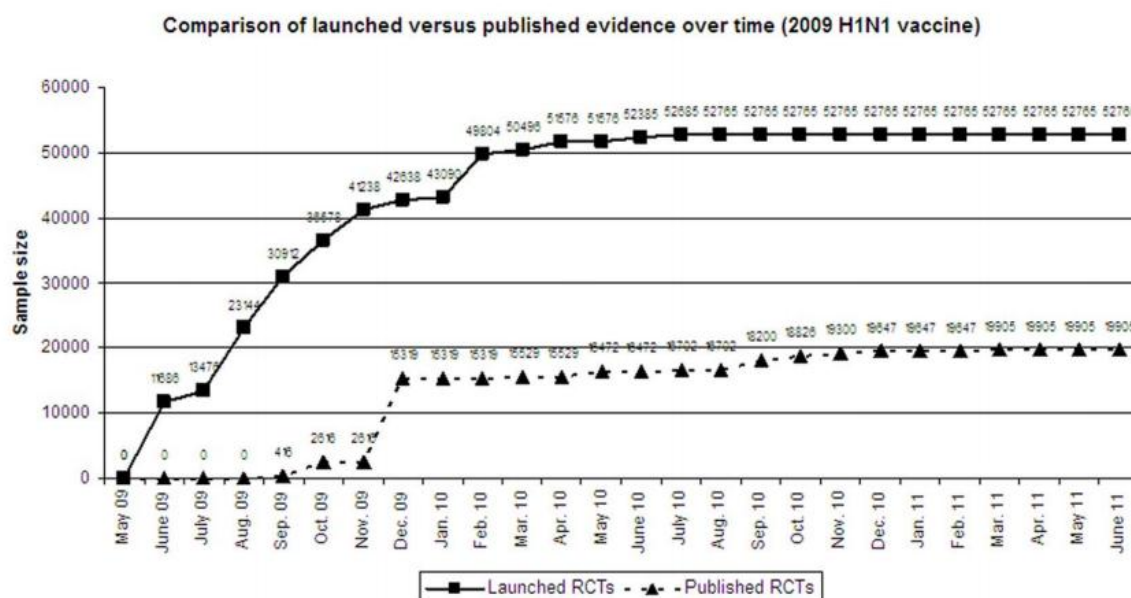

Figure 2. Cumulative sample size in launched and published trials of 2009 H1N1 vaccines over time.
doi:10.1371/journal.pone.0028346.g002

Figure: publication delay, and non-publication, of trials on H1N1 vaccines. Ioannidis et al, 2011.

This is comparable to the poor performance on trial publication throughout medicine, and reflects widespread ongoing problems around access to results. A 2014 systematic review and meta-analysis of 39 cohort studies on publication bias found that on average only half of all trials approved by an ethics committee or posted on a registry went on to be published[24]. A 2013 cross sectional study of large trials (n>500) currently on clinicaltrials.gov found that 29% remained unpublished.[25] A 2015 cohort study on results reporting to clinicaltrials.gov found that only one in five trials complied with US legislation requiring results within 12 months[26], with no penalties issued to date. The 2015 WHO statement on public disclosure of clinical trial results aims to address this issue, requiring clinical trial results to be shared online and submitted for publication within 12 months of study completion, emphasising the benefits of sharing research data, and establishing greater access to primary datasets as an important principle. [27]

Reasons for publication delay highlighted across the existing literature on medicine in general include:
- Time taken by authors to prepare, to write and submit their papers.
- Tendency to first submit results to high profile journals.
- Time taken by journals to review and make decisions about publication
- Time taken to complete the publication process
- Academic rewards favouring journal publications rather than online results sharing
- Tendency to regard "negative" findings as less interesting, or harder to publish.

1.8 Issues with genome data sharing

There are also issues around viral genome sharing. In some respects the challenges here are smaller: in technical terms, the data is inherently somewhat more readily commensurable, and presents fewer issues around confidentiality and identifiability. However other challenges persist: there are intellectual property concerns, and reciprocality concerns, around viral genome; and there are also academic reciprocity issues. It is possible that these challenges have often been overcome in genome data sharing, in part, because there is also a heritage in this field more broadly (eg Human Genome Project) of rapid open data sharing. This heritage derives, at least in part, from the early recognition of the importance of the impending issue, and the conscious direction of funds to promote openess.

However there are also clear shortcomings in this field, as set out by Nathan Yozwiak and colleagues this year[28]. After an initial wave of genome sharing there was "three months of stasis, during which no new virus sequence information was made public.... Some genomes are known to have been generated during this time from patients treated in the United States. The number is likely to have been much larger: thousands of samples were transferred to researchers' freezers across the world. In an increasingly connected world, rapid sequencing, combined with new ways to collect clinical and epidemiological data, could transform our response to outbreaks. But the power of these potentially massive data sets to combat epidemics will be realized only if the data are shared as widely and as quickly as possible. Currently, no good guidelines exist to ensure that this happens." The figure below again demonstrates data sharing, or the lack thereof, alongside disease incidence. As with the graphs of publication delay above, this figure shows that the genome data is not available at the time when it is needed most, and then is shared at a later time when its value and health impact are greatly diminished.
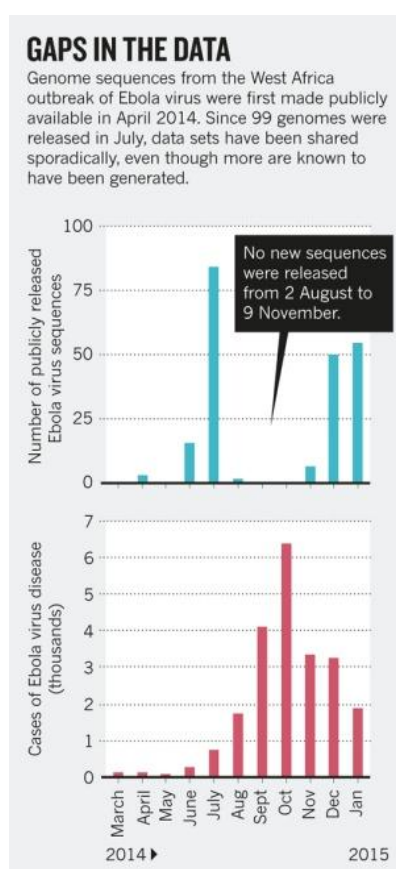

Figure: genome data sharing during the Ebola outbreak, Yozwiak 2015.

1.9 Lack of time

Some interviewees commented that having the time to process the data are as big problems as the subsequent sharing of results.

- *'In the case of emergencies all resources are very limited. I myself found it sometimes difficult to contact such parties and to deliver the information they wanted from me because I was simply that overwhelmed by the outbreak and all the things that we needed to do'*
- *'There is a lot of data cleaning, checking, queries to be done and then you've got to actually evaluate the data because the thing people want to know is does the drug work? It's important to answer this question as early as possible as opposed to just providing data.'*
- *'The setting of unrealistic timelines is not going to facilitate the sharing of data.'*

While emergency settings present extreme challenges, this is consistent with research on failure to publish results in medicine generally: a 2015 systematic review of 24 studies on reasons for failure to pubish found that "lack of time" was the single most common reason researchers gave for not publishing their results in an academic journal after presenting at a conference[29] [30].

2. Suggested solutions to improve data and results sharing in public health emergencies

2.1 General issues around the process of developing solutions

Interviewees raised several general issues around the process of developing solutions to the problems set out above. Several interviewees talked about the need to **learn from others, and previous experiences**.

- *'We need to revisit strategies that have been tried so that we can build on successes and avoid making the same mistakes.'*
- *'Think about how data is shared in the context of other emergencies, which often do turn into public health emergencies - earthquakes, or tsunamis - and how that type of [data and] response is managed. I think there is a lot that we can learn in health sciences and public health in particular from what has been done in those other contexts.'*
- *'For the time being groups with a proven track record of delivery should help shape the response to outbreaks from data generation, data analysis and public release - we must move away from 'teams' relearning the complexities of this type of work every time there is an outbreak'*

There was also concern about the need **create enduring systems** or cultures for sharing information, that can be used when a new group of individuals come to the same set of problems:

- *'Every time there's a new outbreak and it's a new pathogen, then people that have never had to experience how you do this start learning from the bottom up... and they end up making all the mistakes that we keep making of the last 15 or 20 years'*

Another common theme was the need to **listen to, and involve, the right people**, whether these are international organisations or local communities affected by specific outbreaks.

- *'It would be very important to ensure that the key actors in crises are included in the discussion here as most of the time crises are dealt with by NGOs and other UN agencies, like UNHCR, WFP and UNICEF.'*
- *'At meetings on data sharing, the conversation can be skewed or one sided because all of the key stakeholders sometimes are not there.'*
- *'People have very different perspectives and so we need to listen and try to understand what those perspectives are. We need to have an open and honest conversation: "what are you really worried about in terms of sharing data?"'*
- *'Ensure that documents are translated into local languages and explained.'*
- *'Working with the communities to be able to help them understand how you want to use the information being gained from their citizens to be able to maximally help their other citizens.'*

Several interviewees raised the need for **clear positive public commitments** from individuals and organisations in leadership roles.
- *'Statements from high up in national government are very important, and then they are empowered to make that work all the way down through the various chains of command that governments have'*
- *'WHO needs to be more robust in insisting that data is shared publically for the global good rather than distributed internally.'*

And the need for **capacity building** and sustainability in general:

- *'We need to invest in capacity building so that we have the resources to respond efficiently and effectively in public health emergencies.'*

2.2 Regulatory frameworks

Many of the interviewees described the need for an internationally agreed framework and for better systems to be in place, prior to public health emergences.

- *'We need to move towards an international framework for data sharing so that we can be responsive and we need to anchor things in good data sharing principles.'*
- *'Everybody needs to agree on a set of common principles about how we are going to share data together.'*
- *'There really is a need for some move towards a global consensus on governance of data sharing so that various players, not only the public, but certainly even the researchers who collect the data in the first place are protected and fairly treated.'*
- *'I think it is important that there's moves to at least the first round of an international position statement, which will no doubt have to be revised as the field unfolds, but I think it's really timely and very important and it's certainly a big public health issue.'*
- *'MoHs under the direction of the WHO should pre-sign up to release of pathogen genome data, defocused geolocation and date of sampling. Better still this should be linked to line-list epi data — this should be enshrined in SOPs and MOUs'*
- *'We should consider different regulations for surveillance and response regimes and not use the existing regulations that govern research – in some countries data can be collected under infectious disease law and analyses can be conducted without going through usual research regulations.'*

Chatham House, an international affairs thinktank with considerable expertise on diplomacy, is running a series of workshops on improving data sharing in emergencies (which began in 2014 and will report in 2016). Their preliminary report is notably sceptical on legislative harmonisation and favours other approaches, specifically a global data governance and ethical framework.

2.3 Knowledge curation

Many interviewees expressed a need for better technical systems for knowledge curation and data sharing, with various models implicitly proposed, ranging from formal and structured to informal. Specific technical details of knowledge and data sharing platforms were rarely covered in detail by interviewees.

- *'We need to build databases where all data are entered in a uniform way which can be populated when outbreaks occur and are available worldwide.'*
- *'Information is often shared informally and we can make better use of informal networks.'*
- *'Online conferences where emerging data can be presented should be utilized.'*
- *'Simple ways of reporting the data back to MoH and on the ground epi teams should be enshrined on SOPs and resource'*
- *'These databases should be accessible for skilled individuals who won't use it for personal or political gain, people who are skilled in the art of doing it and who have made a commitment to use the data to help the public health'*

WHO and CDC already publish public health reports and updates with rapid turnaround times.[31] Several interviewees raised the opportunities of using WHO as a central source of information more generally.

- *'At the beginning of 2015, I have the feeling that WHO became a more important point of sharing information to third parties, there were information points set up which were helpful for all the companies who had questions – this was very important to have a central point where information could be delivered and where I could refer people to'.*
- *'Who has the legitimacy to do this kind of work and who has the technical capabilities? The technical expertise does not lie within WHO but the mandate to do it does so it needs to be a partnership.'*

A rapid overview of existing formal knowledge exchange and data sharing platforms is provided below. This overview is likely to be incomplete and we would welcome suggestions of additional resources.

EbolaClinicalTrials.org is a knowledge exchange platform for methods and resources on clinical trials of interventions for Ebola virus disease. Investigators have already formed informal networks to share data and preliminary results with colleagues from other countries.

The **WWARN Worldwide Antimalarial Resistance Network** is a collaborative initiative with a wide remit including data sharing, knowledge curation, professional networking, and online training, with a large user base.

The **Ebola Virus Disease Data Sharing Platform Initiative** is a project currently being launched, developed by WWARN personnel, to share clinical, laboratory and epidemiological data on Ebola.

GISAID, the Global Initiative on Sharing Avian Influenza Data, is a global data sharing and collaboration platform. It is run on the principle that genome data, along with associated clinical and epidemiological data, should be openly shared, and then collaboratively analysed and published. GISAID is a longstanding platform with a large user base who have committed to these principles. However there have been concerns about intellectual property issues, for example that the data may be subject to commercial exploitation for the development of vaccines without reciprocation, and so some have sought to protect their data from such exploitation before posting.

ISARIC, the International Severe Acute Respiratory and Emerging Infection Consortium is a global initiative aiming to develop and share open access protocols and harmonised data-sharing processes for emerging diseases that may turn into epidemics or pandemics. It gathers over 70 networks and individuals involved in research related to diseases such as bird flu (H5N1), swine flu (H1N1) and SARS. They aim to produce "Protocols, Informed Consent Forms, Case Report Forms, Standard Operating Procedures, training materials and other research tools are available to support implementation of the research." Under protocols and data tools the ISARIC site currently lists one protocol (co-produced by ISARIC with WHO, a standardised observational study for rapid coordinated clinical investigation of patients with severe infections); and two data tools (a 14 page long Case Record Form for Severe Acute Respiratory Infection, and similar for Ebola). There is also a link to a centralised online data management system for approved users collaborating on ISARIC work, at https://www.cliresdms.org/

Interviewees discussed some of the positive characteristics of initiatives such as those listed above. For example:

- *'In terms of success, I think one of the reasons they have been successful is that all investigators who contribute data to the initiative are recognized in all publications, it is housed at a neutral*

entity and there is a discussion and communication group that goes along with it. They also provide a clearinghouse of other information.'

There are also multiple initiatives specifically for genome data sharing. **NextFlu** take real time virus genome sequences and limited meta data and make that information publicly available. Much data on the Ebola virus was rapidly made available on **virological.org**. The Global Influenza Surveillance Network **GISN** facilitated the rapid sharing and analysis of virological specimens s from early cases of H1N1 pandemic influenza. **GenBank,** the NIH genetic sequence database, is an annotated collection of all publicly available DNA sequences from over 100,000 organisms to date. The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. The National Institutes of Health (NIH) Genomic Data Sharing (**GDS**) Policy promotes sharing for research purposes of genomic data generated by NIH-funded research. [32] Under this policy investigators are expected to budget for resources needed to prepare and submit data to appropriate repositories along with timelines for data submission and release.

There are also numerous policy initiatives around data sharing in science more broadly, outside of emergency settings, and in some cases outside of health. These include several initiatives by major research funders, and longstanding work by the **International Council for Science**, which has two operations aiming to facilitate and promote data sharing in science: CODATA was established in 1966 as an interdisciplinary committee to develop policy and norms on data sharing in science; the World Data System aims to facilitate data sharing and promote the adoption of standards.

2.4 Improving publication processes

Positive suggestions around expedited journal publication were repeatedly raised by interviewees.

- *'Provisional research should be published and subjected to ongoing peer review as more data becomes available.'*
- *'We should provide results to governments and the people who need the information before scientific publication.'*

There are several journal initiatives already ongoing around expedited publication. Examples include:

- The NEJM consider fast track requests and authors should hear back within 36 hours, and a decision on publication within 2 to 3 weeks [33]
- PLoS Currents: Influenza commits to rapidly screens articles by leading researchers in the field.
- The Lancet offers rapid review ('For research papers that are judged to warrant fast dissemination, which will usually be randomised controlled trials, The Lancet will publish a peer-reviewed manuscript within 4 weeks of receipt').[34]
- The BMJ published online Influenza resources and offered rapid review.
- Many other journals have offered expedited review for papers on public health emergencies.
- There are longstanding discussions on the benefits of protocol review for trials, leading to definitive acceptance or rejection prior to the main results being available.

There are also novel publishing models. F1000Research will post submissions online almost immediately, with open peer-review after publication. Articles are tracked, so readers can be notified when expert referee reports are made available.[35] Early in the Ebola outbreak F1000Research made an explicit offer of rapid publication of all Ebola research, including a telephone line that researchers or clinicians at remote locations could call, to dictate reports direct to the F1000R office. This offer was

accompanied by an endorsement from Peter Piot, the Director of LSHTM and co-discoverer of Ebola.[36] Other related initiatives such as *Nature Precedings -* modelled on preprint servers in other disciplines - have previously failed to sustain their activities.[37]

The ability of journal editors to distinguish important evidence during a pandemic may be limited, and superficially uninteresting findings may be of importance to someone somewhere, with sufficiently rapid dissemination, if only to avoid duplication of research effort.

2.5 Changing academic reward structures

Various interviewees discussed a need to change the way academic rewards are currently structured, away from journal publication to data sharing and its subsequent impact.

- *'You really need a culture change at the bottom where people that really are generating this data are actively empowered by their heads of organizations to make that data rapidly available (bottom up)'*
- *'There needs to be some sort of recognition or reward mechanism that actually rewards people in being altruistic and acknowledges them appropriately'*

This is reflected in the wider literature. For example, a Lancet series on "reducing waste in research" similarly proposed that reward structures in academia should target effective data sharing by researchers ('academic institutions and funders should reward investigators who fully disseminate their research protocols, reports, and participant-level datasets').[38]

2.6 Additional suggestions from CEBM

CEBM has also been asked to raise suggestions on further interventions that may improve data and results sharing. Although CEBM does not have staff working on public health emergencies it does have experience on knowledge synthesis, knowledge translation, and addressing shortcomings in the information architecture around evidence based medicine. Further to the solutions discussed above, we would add:

1. It is a worthwhile exercise to attempt to define the ideal regulatory frameworks, and data sharing platforms, even if these are not attainable. These ideals can act as a target, to facilitate gap analysis, and for real-world use by a smaller proportion of all actors as a live demonstration of the benefits of best practice.

2. It may be worth developing a framework for centralised knowledge curation, in collaboration with information management experts, to be staffed by health information specialists available on a "surge" basis during public health emergencies. This need not be the only option for knowledge pooling and distribution, but it would be wise to exploit the existing knowledge base on curating information, especially health information.

3. To address delayed publication or non-publication of trials and other research during an emergency, a simple ongoing public audit identifying and naming missing studies is better than producing a publication - itself years after the emergency has passed - documenting the proportion of delayed publications during the emergency. Such an audit would consist of a simple table, setting out

what studies have commenced, which have completed, which have shared results, which have failed to do so, and the number of days since study completion without results being shared for each study. This can be used for live feedback, and public accountability on unpublished results may help to drive up standards.

4. It may be worth reviewing whether legally robust but clear and comprehensible "off the shelf" data sharing licenses, following the Creative Commons model, can help to address the concerns of data holders about releasing data more widely. Such a model was alluded to, in principle at least, by one interviewee ('there's a real need for explaining and setting up with standard operating procedures, terms of reference, canned if you like permission structures with various ministries of health exactly what can and should be expected if an outbreak appears in your jurisdiction'). Each public health emergency can mobilise a different cast of players, with different norms, experiences, expectations, and needs. Negotiating complex agreements from scratch can be unwieldy. Ready made licenses for data sharing could set out the permissions, attribution, audience, and strengths/weaknesses of each option. These off-the-shelf agreements could be provided alongside examples of what others have chosen before, to use social norms as advocacy, or to give reference and guidance on what others in similar positions have done.

5. Journals remain a key vehicle for communicating research results, but inevitably present delays. In light of this, various novel approaches may be valuable. For example: journals could rapidly pass research on to a different journal when it is rejected on grounds of interest or quality, and share peer review information to avoid duplication of effort or sequential delays.

References

1 Yozwiak NL, Schaffner SF, Sabeti PC. Data sharing: Make outbreak research open access. Nature. 2015 Feb 26;518(7540):477-9.

2 Bull, Susan, Phaik Yeong Cheah, Spencer Denny, Irene Jao, Vicki Marsh, Laura Merson, Neena Shah More, et al. 'Best Practices for Ethical Sharing of Individual-Level Health Research Data From Low- and Middle-Income Settings.' Journal of Empirical Research on Human Research Ethics 10, no. 3 (1 July 2015): 302–13. doi:10.1177/1556264615594606.

3 Vallance P, Chalmers I. Secure use of individual patient data from clinical trials. Lancet.2013;382(9898):1073–1074.

4 Appendix B, "Concepts and Methods for De-identifying Clinical Trial Data" in: 'Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.' Institute of Medicine. Accessed 26 August 2015. http://iom.nationalacademies.org/Reports/2015/Sharing-Clinical-Trial-Data.aspx

5Bull, Susan, Phaik Yeong Cheah, Spencer Denny, Irene Jao, Vicki Marsh, Laura Merson, Neena Shah More, et al. 'Best Practices for Ethical Sharing of Individual-Level Health Research Data From Low- and Middle-Income Settings.' Journal of Empirical Research on Human Research Ethics 10, no. 3 (1 July 2015): 302–13. doi:10.1177/1556264615594606.

6 Public Health Collection, Use, Sharing, and Protection of Information. http://www.astho.org/Programs/Preparedness/Public-Health-Emergency-Law/Public-Health-and-Information-Sharing-Toolkit/Collection-Use-Sharing-and-Protection-Issue-Brief/

7 Jussi Sane and Michael Edelstein Centre on Global Health Security Overcoming Barriers to Data Sharing in Public Health: A Global Perspective. April 2015 Chatham House. OvercomingBarriersDataSharingPublicHealthSaneEdelstein.pdf

8 European Commission. 'Lessons Learned from the 2011 Outbreak of Shiga Toxin-Producing Escherichia Coli (STEC) O104:H4 in Sprouted Seeds', 2011. http://ec.europa.eu/food/food/biosafety/salmonella/docs/cswd_lessons_learned_en.pdf.

9 The Ebola emergency perspectives on information management and mapping responses - Issue 64 - Humanitarian Exchange Magazine - Humanitarian Practice Network. www.odihpn.org/humanitarian-exchange-magazine/issue-64/the-ebola-emergency-perspectives-on-information-management-and-mapping-responses

10 Sylvie Briand. Challenges of global surveillance during an influenza pandemic http://www.who.int/influenza/surveillance_monitoring/Challenges_global_surveillance.pdf

11 Roheeni Saxena. Ebola created a public health emergency—and we weren't ready for it. Arstechnica. http://arstechnica.co.uk/science/2015/07/ebola-created-a-public-health-emergency-and-we-werent-ready-for-it/

12 Huang Y. THE SARS EPIDEMIC AND ITS AFTERMATH IN CHINA: A POLITICAL PERSPECTIVE. In: Institute of Medicine (US) Forum on Microbial Threats; Knobler S, Mahmoud A, Lemon S, et al., editors. Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary. Washington (DC): National Academies Press (US); 2004. Available from: http://www.ncbi.nlm.nih.gov/books/NBK92479/

13 BBC (British Broadcasting Corporation) Zimbabwe cholera is over—Mugabe. Dec 11, 2008b. [accessed January 27, 2009]. http://news.bbc.co.uk/2/hi/africa/7777178.stm.

14 Institute of Medicine (US) Forum on Microbial Threats. Ethical and Legal Considerations in Mitigating Pandemic Disease: Workshop Summary. Washington (DC): National Academies Press (US); 2007. 1, Learning from Pandemics Past. Available from: http://www.ncbi.nlm.nih.gov/books/NBK54171/

15 Institute of Medicine (US) Forum on Microbial Threats. Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary. Edited by Stacey Knobler, Adel Mahmoud, Stanley Lemon, Alison Mack, Laura Sivitz, and Katherine Oberholtzer. The National Academies Collection: Reports Funded by National Institutes of Health. Washington (DC): National Academies Press (US), 2004. http://www.ncbi.nlm.nih.gov/books/NBK92462/.

16 International Health Regulations (2005) Second edition. World Health Organization http://www.who.int/ihr/9789241596664/en/

17 Langat P, Pisartchik D, Silva DS, Bernard C, Olsen K, Smith M, Sahni S, Upshur REG (2011) Is There a Duty to Share? The Ethics of Sharing Research Data in the Context of Public Health Emergencies. Public Health Ethics 4(1): 4-11.

18 Merson, Laura, Tran Viet Phong, Le Nguyen Thanh Nhan, Nguyen Thanh Dung, Ta Thi Dieu Ngan, Nguyen Van Kinh, Michael Parker, and Susan Bull. 'Trust, Respect, and Reciprocity Informing Culturally Appropriate Data-Sharing Practice in Vietnam.' Journal of Empirical Research on Human Research Ethics 10, no. 3 (1 July 2015): 251–63. doi:10.1177/1556264615592387.

19 Bull, Susan, Phaik Yeong Cheah, Spencer Denny, Irene Jao, Vicki Marsh, Laura Merson, Neena Shah More, et al. 'Best Practices for Ethical Sharing of Individual-Level Health Research Data From Low- and Middle-Income Settings.' Journal of Empirical Research on Human Research Ethics 10, no. 3 (1 July 2015): 302–13. doi:10.1177/1556264615594606.

20 Bull, Susan, Nia Roberts, and Michael Parker. 'Views of Ethical Best Practices in Sharing Individual-Level Data From Medical and Public Health Research A Systematic Scoping Review.' Journal of Empirical Research on Human Research Ethics 10, no. 3 (1 July 2015): 225–38. doi:10.1177/1556264615594767.

21 Xing W, Hejblum G, Leung GM, Valleron AJ. Anatomy of the epidemiological literature on the 2003 SARS outbreaks in Hong Kong and Toronto: a time-stratified review. PLoS Med. 2010 May 4;7(5):e1000272. doi: 10.1371/journal.pmed.1000272. Review.

22 Ioannidis JP, Manzoli L, De Vito C, D'Addario M, Villari P. Publication delay of randomized trials on 2009 influenza A (H1N1) vaccination. PLoS One. 2011;6(12):e28346. doi: 10.1371/journal.pone.0028346. Epub 2011 Dec 2.

23 Ott M, Shaw SF, Danila RN, Lynfield R. Lessons learned from the 1918-1919 influenza pandemic in Minneapolis and St. Paul, Minnesota. Public Health Rep. 2007 Nov-Dec;122(6):803-10. No abstract available.

24 Schmucker, Christine, Lisa K. Schell, Susan Portalupi, Patrick Oeller, Laura Cabrera, Dirk Bassler, Guido Schwarzer, et al. 'Extent of Non-Publication in Cohorts of Studies Approved by Research Ethics

Committees or Included in Trial Registries.' PLoS ONE 9, no. 12 (23 December 2014): e114023. doi:10.1371/journal.pone.0114023.

25 Jones, Christopher W., Lara Handler, Karen E. Crowell, Lukas G. Keil, Mark A. Weaver, and Timothy F. Platts-Mills. 'Non-Publication of Large Randomized Clinical Trials: Cross Sectional Analysis.' BMJ 347 (29 October 2013): f6104. doi:10.1136/bmj.f6104.

26 Anderson, Monique L., Karen Chiswell, Eric D. Peterson, Asba Tasneem, James Topping, and Robert M. Califf. 'Compliance with Results Reporting at ClinicalTrials.gov.' New England Journal of Medicine 372, no. 11 (12 March 2015): 1031–39. doi:10.1056/NEJMsa1409364.

27 WHO Statement on Public Disclosure of Clinical Trial Results. http://www.who.int/ictrp/results/reporting/en/

28 Yozwiak, Nathan L., Stephen F. Schaffner, and Pardis C. Sabeti. 'Data Sharing: Make Outbreak Research Open Access.' Nature 518, no. 7540 (25 February 2015): 477–79. doi:10.1038/518477a.

29 Researchers "lack time" to publish results, new review finds. Alltrials news 24th February 2015 http://www.alltrials.net/news/researchers-lack-time-to-publish-results-new-review-finds/

30 Scherer, Roberta W. et al. Authors report lack of time as main reason for unpublished research presented at biomedical conferences: a systematic review. Journal of Clinical Epidemiology , Volume 68 , Issue 7 , 803 – 810

31 Sylvia Briand,. Challenges of global surveillance during an influenza pandemic. WHO http://www.who.int/influenza/surveillance_monitoring/Challenges_global_surveillance.pdf

32 NIH Genomic Data Sharing Policy. Notice Number: NOT-OD-14-124 http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html

33 PRESUBMISSION INQUIRIES/FAST TRACK. NEJM .http://www.nejm.org/page/author-center/manuscript-submission

34 The Lancet fast track review. http://www.thelancet.com/lancet-information-for-authors/fast-track. Accessed 26 August 2015.

35 http://f1000r.es/4ot] F1000Research 2014, 3:269 (doi: 10.12688/f1000research.5685.1)

36 http://blog.f1000research.com/2014/11/12/beat-ebola-with-better-research-sharing-says-discoverer-of-virus/

37 Nature Precedings. http://precedings.nature.com/ Accessed 26 August 2015.

38 Chan, An-Wen et al.Increasing value and reducing waste: addressing inaccessible research. The Lancet , Volume 383 , Issue 9913 , 257 - 266